CSPA

Common Statistical Production Architecture

State of the art of current sharing activities between NSIs and statistical institutions

Carlo Vaccari
Istat (vaccari@istat.it)

Index

- CORA survey on architecture, tools and sharing
- SAB repository
- Other documents

The Questionnaire

One questionnaire was designed to ask NSIs about Common Reference Architecture requirements

and shared software.

ITDG in 2009 meeting gave support to the survey

Outputs:

- Set of requirements for CORA
- State of the art analysis
- Documents on EA
- Software tools available for sharing

The Questionnaire

The questionnaire was divided into 7 sections:

1. Introduction - name and size of NSIs
2. Statistical BPM usage (GSBPM usage)
3. GSBPM sub-process for phase 5 analysis
4. Enterprise Architecture adoption
5. Tools: sharable and open source (usage, policies)
6. State of the art - documentation
7. Comments

The questionnaire is here

Shared/sharable tools

Then in the questionnaire characteristics were asked either for shared tools and for sharable tools

Results from this questionnaire has been published under the Software Sharing part of the MSIS (Management of Statistical Information Systems) wiki here

Cooperation with SAB (Sharing Advisory Board) a subsidiary of the Conference of European Statisticians (CES) with those duties

The Survey

Target: 42 NSIs = EU + candidates + EFTA + others MSIS

The total response rate of the survey was 79% (33 NSIs replied to the survey).

The response rate within the ESS was 80% (21 EU member states and 3 EFTA countries).

The organisation size ranged from 74 to 5600 employees with an average size of 1202.

A total of 66 individuals were involved in filling out the 33 questionnaires.

Results: BPM

42% (14) of the organisations have an official statistical business process model, many of them used the BPM to restructure their organizations and to guide IT developments

Some of them had already developed a mapping between their BPM and GSBPM

Results: phase 5 GSBPM (Process)

Choose one phase 5 sub-phase between

5.1 Integrate data	3
5.2 Classify and code	7
5.3 Review, validate and edit	12
5.4 Impute	7
5.5 Derive new variables & statistical units	1
5.6 Calculate weights	6
5.7 Calculate aggregates	6
5.8 Finalize data files	2

Results: phase 5 GSBPM (Process)

For the choosen sub-phase:

- standardization degree: current and potential
- data interface currently used (DB 17/48, ETL 8/48, Standard file 16/48, …)
- plans for improvement: 29/44
- reasons for choosing this subprocess: 32/44 familiar, 25/44 working well, 10/44 challenging

Results: Enterprise Architecture

70% have an IT strategy document

52% have principles (business, architecture etc) intended to guide IT decision making

12 % have adopted/ implemented an Enterprise Architecture Framework (EAF)

27% are using a large variety of EA tools

15% are testing EA tools.

30% have architectural roles staffed (formally or informally) in the EA

Results: Tools

85% use shared software tools i.e. tools developed by producers of official statistics and currently shared with other official statistics producers

42% have candidate software tools i.e. tools which are not currently shared but could be further developed to serve a wider community of producers of official statistics

27% have provided ideas for developing new shared software tools

27% have an official policy on the use of OSS

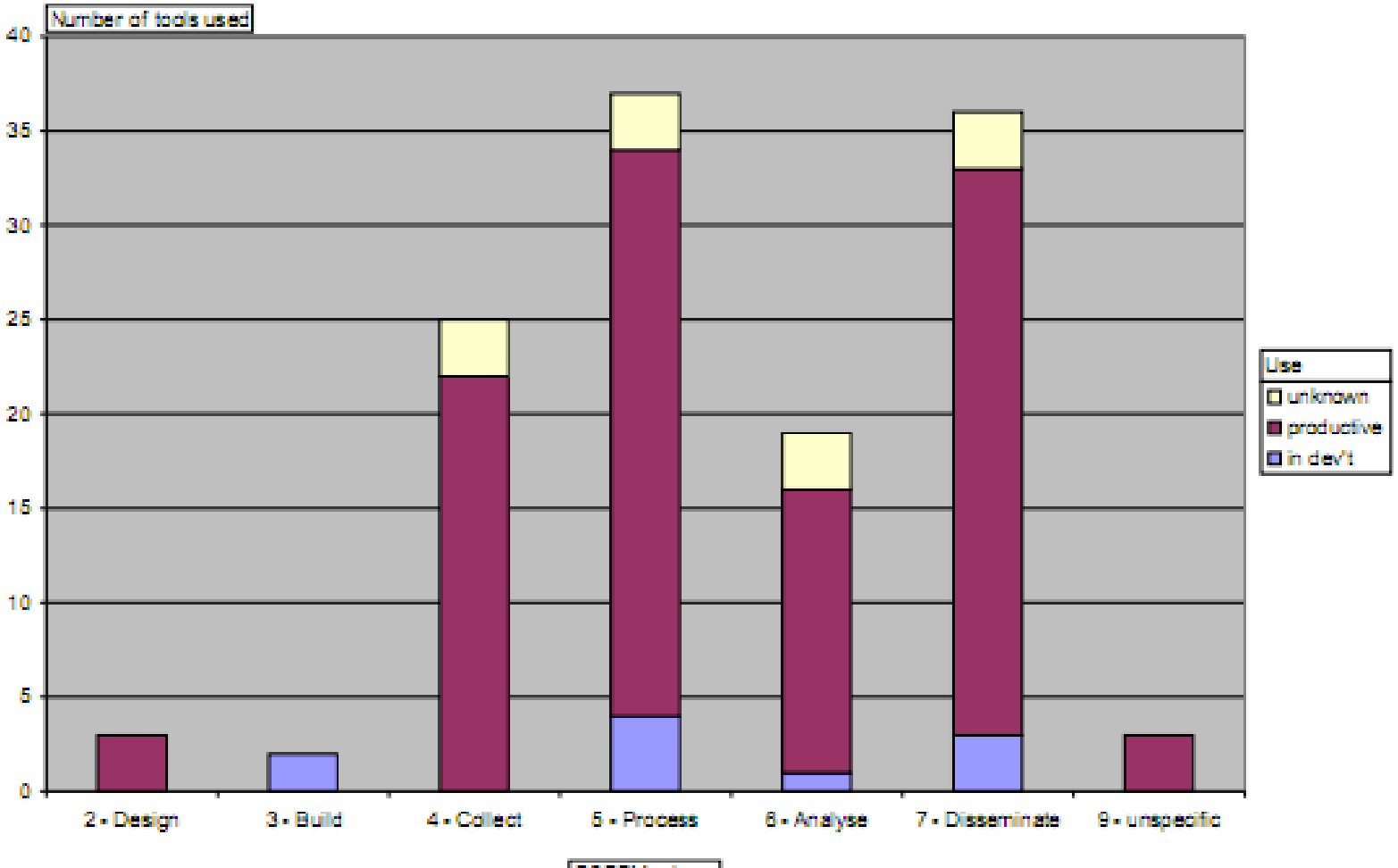
55% use OSS in the statistical production process

Results: additional documents

More than 90 documents or links about

- Organisation structure
- Official statistical business process model and mapping with the GSBPM
- IT strategy document
- Principles to guide IT decision making
- Enterprise Architecture Framework (EAF) and EA tools
- Policy on the use of Open Source Software
- State of the art documentation

Software tools shared and sharable



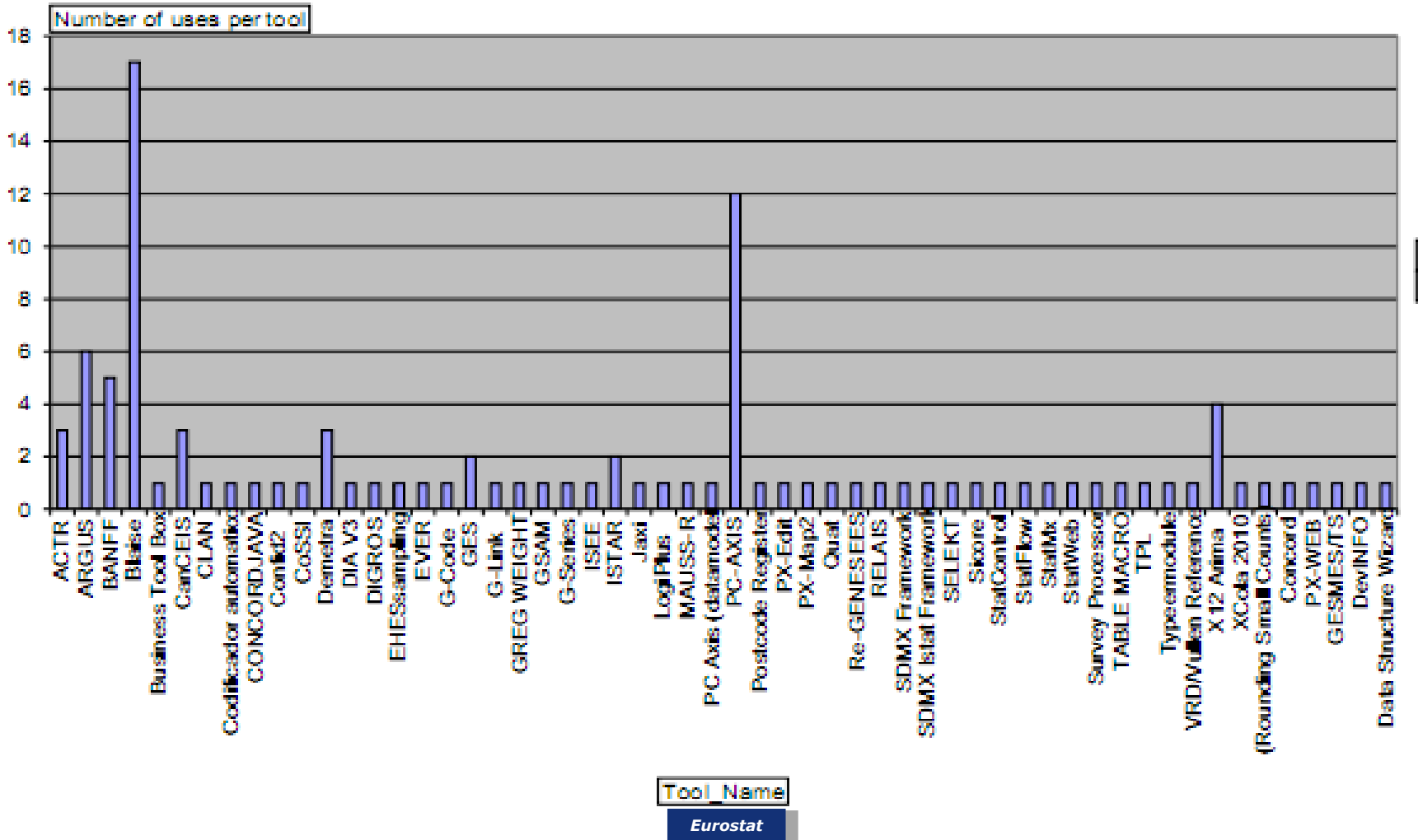
Software tools shared and sharable

Software tools specifically developed by national statistics institutes for their statistical purposes, made available for other NSIs under different conditions. Commercial tools are not included in this survey

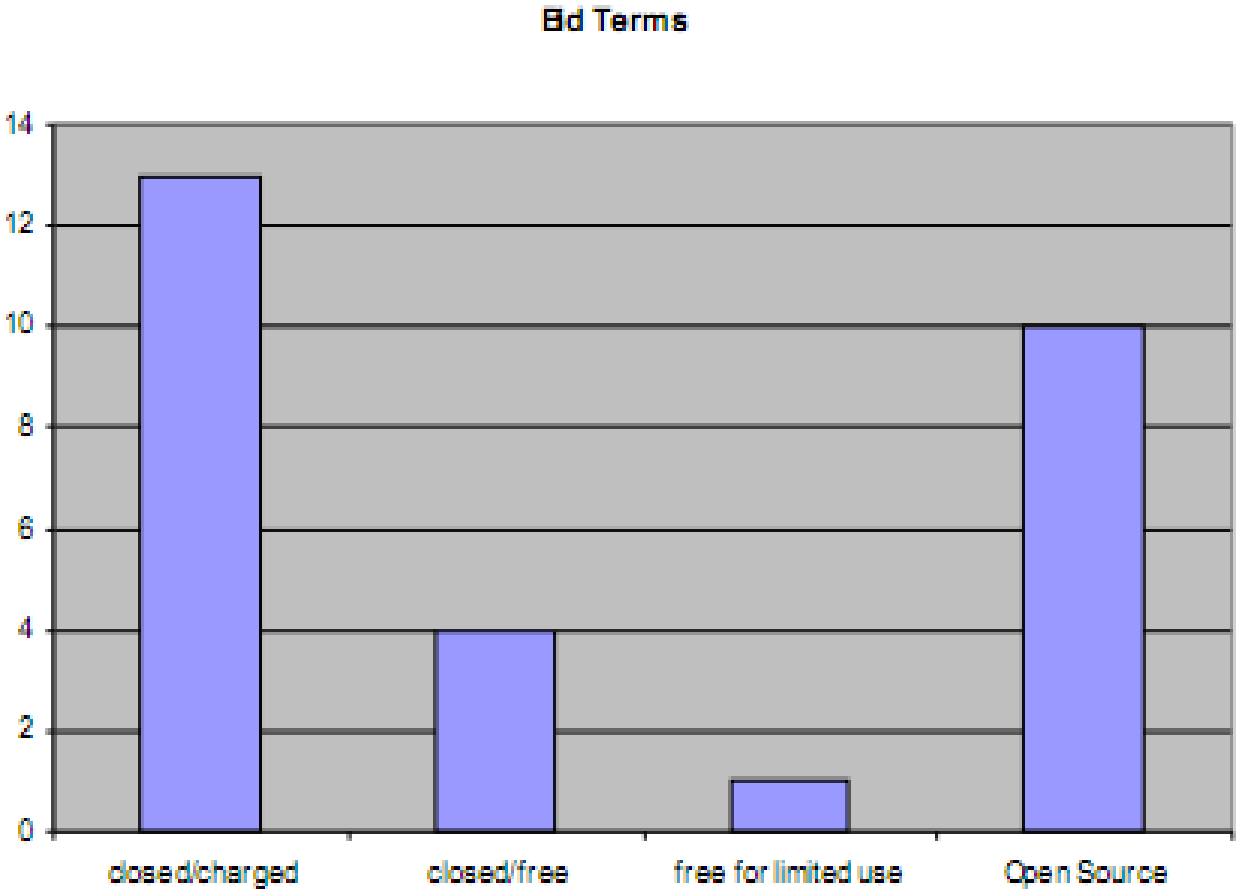
As a first result, this survey shows that most of the software is targeted for the “Process” and “Disseminate” phases of the GSBPM, followed by “Collect” and “Analyse” phases

Very few tools for the “Build” and “Design” phases; probably because this is a domain where non-statistical organisations and commercial providers are active

Software tools shared and sharable



Software tools shared and sharable



Software tools shared and sharable

Number of tools used	Use			
GSBPM_phase	in dev't	productive	unknown	Total
2 - Design		3		3
3 - Build	2			2
4 - Collect		22	3	25
5 - Process	4	30	3	37
6 - Analyse	1	15	3	19
7 - Disseminate	3	30	3	36
9 - unspecific		3		3
Total	10	103	12	125

Number of times different tools are used

Software tools shared and sharable

Everything stored in Sharing Advisory Board

Now updated: Software Inventory

See list, examples and data available

Connection with HLG-BAS and GSIM

Recent developments

In 2012 compliance with the HLG vision. Software must:

- Support input and output of data (and metadata) in "open" formats, e.g. CSV, DDI, XML, SDMX
- Be capable of being used in a "plug and play" architecture based on the SOA model. Key requirements for the HLG vision are:
  - Having clearly defined interfaces to run the software
  - Having clearly defined outputs from running the software - both during run-time and at the end
  - Using the SOA concept of appropriate granularity, based on the sub-process level of the GSBPM
  - Designing and implementing for re-use by others.

Comply with the
Guidelines on Developing Multilingual Applications for Official Statistics

Other links

- Newsletter
- Prerequisite for improved collaboration: link
- Cooperation models for software developments
- Software for Data Integration: Record Linkage
- Collaboration projects